ORIGINAL ARTICLE

Precise detection of chromosomal translocation or inversion breakpoints by whole-genome sequencing

Toshifumi Suzuki[1,2], Yoshinori Tsurusaki[1], Mitsuko Nakashima[1], Noriko Miyake[1], Hirotomo Saitsu[1], Satoru Takeda[2] and Naomichi Matsumoto[1]

Structural variations (SVs), including translocations, inversions, deletions and duplications, are potentially associated with Mendelian diseases and contiguous gene syndromes. Determination of SV-related breakpoints at the nucleotide level is important to reveal the genetic causes for diseases. Whole-genome sequencing (WGS) by next-generation sequencers is expected to determine structural abnormalities more directly and efficiently than conventional methods. In this study, 14 SVs (9 balanced translocations, 1 inversion and 4 microdeletions) in 9 patients were analyzed by WGS with a shallow (5×) to moderate read coverage (20×). Among 28 breakpoints (as each SV has two breakpoints), 19 SV breakpoints had been determined previously at the nucleotide level by any other methods and 9 were uncharacterized. BreakDancer and Integrative Genomics Viewer determined 20 breakpoints (16 translocation, 2 inversion and 2 deletion breakpoints), but did not detect 8 breakpoints (2 translocation and 6 deletion breakpoints). These data indicate the efficacy of WGS for the precise determination of translocation and inversion breakpoints.

INTRODUCTION

Structural variations (SVs), including translocations, inversions, deletions and duplications, potentially lead to human genetic diseases arising from disruption and dosage changes of functionally important genes.[1] In particular, apparently balanced chromosomal rearrangements have been frequently associated with human diseases, such as premature ovarian failure, Sotos syndrome, Peters anomaly, testicular atrophy, Mowat–Wilson syndrome, developmental delay and intellectual disability.[2–6] The incidence of apparently balanced chromosomal rearrangements is in the range of 1/500–1/625.[7] Precise structural analysis of SVs and their breakpoints may lead to identification of the genetic causes of such diseases. The conventional methods to determine SV breakpoints include fluorescence in situ hybridization (FISH) using bacterial artificial chromosome clones, Southern blot hybridization and inverse PCR or long-range PCR, which are laborious, time-consuming and have limited success rates.[8] Recently, whole-genome sequencing (WGS) using next-generation sequencers has provided a new avenue for SV analysis.[2,7–9] However, accurate detection of SV breakpoints using WGS has not been fully established. In this study, WGS was used to analyze 9 patients having 14 SVs. As each SV has 2 breakpoints, 28 SV breakpoints were analyzed. Among them, 19 SVs had already been determined by conventional methods in our previous studies[6,10–13] and used as a training set, and 9 other uncharacterized breakpoints were analyzed. The purpose of this study is to investigate the chromosomal breakpoint of the patients in whom G-banded karyotyping was already performed. The results of WGS analysis of these patients are presented.

MATERIALS AND METHODS

Subjects

Nine patients, including eight who were reported previously,[6,10–14] were included in this study. G-banded karyotyping was performed for all patients (Table 1). The 9 patients possess a total of 14 SVs (9 translocations, 1 inversion and 4 microdeletions) (Tables 1 and 2). As each SV event involves two breakpoints, a total of 28 SV breakpoints are the targets of this study. Among 28 SV breakpoints, 19 were previously determined at the nucleotide level by conventional methods[6,10–14] and used as a training set (Tables 1 and 2). Peripheral blood samples were collected from all patients after obtaining written informed consent. Genomic DNA was extracted from leukocytes using the QuickGene-610L DNA extraction system (Fujifilm, Tokyo, Japan) according to the manufacturer's instruction. The institutional review board of Yokohama City University School of Medicine approved the study.

Whole-genome sequencing

Briefly, 1 µg of genomic DNA with each sample was shared using the Covaris model S2 system (Covaris, Woburn, MA, USA). The target size was 350 bp. DNA was prepared using the TruSeq DNA Sample Prep Kit (Illumina, San Diego, CA, USA) or the TruSeq DNA PCR-Free Sample Prep Kit (Illumina). The HiSeq 2000 or 2500 platform (Illumina) was used to perform WGS with 101-bp paired-end reads. Sequence-control, software real-time analysis and CASAVA software v1.8.2 (Illumina) were used to perform image analysis and base calling.

[1]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan and [2]Department of Obstetrics and Gynecology, Juntendo University Faculty of Medicine, Tokyo, Japan
Correspondence: Professor N Matsumoto, Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama 236-0004, Japan.
E-mail: naomat@yokohama-cu.ac.jp

Table 1 Detection of chromosomal breakpoints by BreakDancer (translocation or inversion)

Patient	Karyotype	Breakpoint detection by other methods (hg19)	Mean coverage (x)	Preparation	PCR duplication (%)	BreakDancer detection	IGV Chimeric read pair counts	IGV Discordant read pair counts	IGV Soft-clipped read counts	References
1	46,XX,t(1;9)(q32;q13)	der(1): 206574512 / der(9): **undetermined**	15.23	PCR-free	1.6	der(1): **negative** / der(9): **negative**	der(1): **0** / der(9): **0**	—	der(1): 1 / der(9): **0**	Saitsu et al.[10]
2	46,XX,t(5;15)(q13.3;q26.1)	der(5): 88300578 / der(15): 98166676	9.18	PCR	7.0	der(5): positive / der(15): positive	der(5): 14 / der(15): 14	—	der(5): **0** / der(15): **0**	Saitsu et al.[11]
3	46,X,t(X;4)(q21.3;p15.2)	der(X): 107436209 / der(4): 12244290	21.92	PCR-free	1.9	der(X): positive / der(4): positive	der(X): 14 / der(4): 12	—	der(X): 5 / der(4): 2	Nishimura-Tadaki et al.[12]
4	46,XX,t(7;15)(q21;q15), t(9;14)(q21;q11.2)	der(7): 96377471 / der(15): 45376277 / der(9): **undetermined** / der(14): **undetermined**	18.49	PCR-free	1.8	der(7): positive / der(15): positive / der(9): positive / der(14): positive	der(7): 18 / der(15): 18 / der(9): 31 / der(14): 31	—	der(7): **0** / der(15): 4 / der(9): 3 / der(14): 5	Saitsu et al.[14]
5	46,X,t(X;2)(q22;p13)	der(X): 107946031 / der(2): 66226306	20.35	PCR-free	2.1	der(X): **negative** / der(2): **negative**	der(X): **0** / der(2): **0**	—	der(X): **0** / der(2): **0**	Nishimura-Tadaki et al.[12]
6	46,X,t(X;4)(q22.1;q12)	der(X): 101594056 / der(4): 57969312	11.69	PCR	45.8	der(X): positive / der(4): positive	der(X): 20 / der(4): 20	—	der(X): **0** / der(4): **0**	Nishimura-Tadaki et al.[12]
7	46,X,t(X;14)(q24;q32.1)	der(X): 117318221 / der(14): 91614197	12.93	PCR	21.3	der(X): positive / der(14): positive	der(X): 24 / der(14): 24	—	der(X): 2 / der(14): **0**	Nishimura-Tadaki et al.[12]
8	46,XX,t(5;8)(q35;q24.1)	**undetermined**	5.95	PCR-free	1.2	der(5): positive / der(8): positive	der(5): 14 / der(8): 14	—	der(5): 2 / der(8): 2	Imaizumi et al.[6] / Kurotaki et al.[13]
9	Inv(5)(p15q13)	**undetermined**	14.37	PCR-free	1.8	5p15: positive / 5q13: positive	—	chr5: 28 (inversion)	5p15: 3 / 5q13: 6	This study

Abbreviation: IGV, Integrative Genomics Viewer.
Bold numbers and text indicate the detection failure.

Table 2 Detection of chromosomal breakpoints by BreakDancer (deletion)

Patient	Breakpoint detection by other methods (hg19)	BreakDancer detection	IGV	
			Discordant read pair counts	Soft-clipped read counts
3	chrX: 107,436,210-107,440,401	chrX: **negative**	chrX: **0**	chrX: **0**
	chr4: 12,237,261-12,244,289	chr4: **negative**	chr4: **0**	chr4: **0**
4	chr7: 96,949,067-97,755,363	chr7: positive	chr7: 9	chr7: 2
	chr15: **undetermined**	chr15: **negative**	chr15: **0**	chr15: **0**

Abbreviation: IGV, Integrative Genomics Viewer.
Bold numbers and text indicate the detection failure.
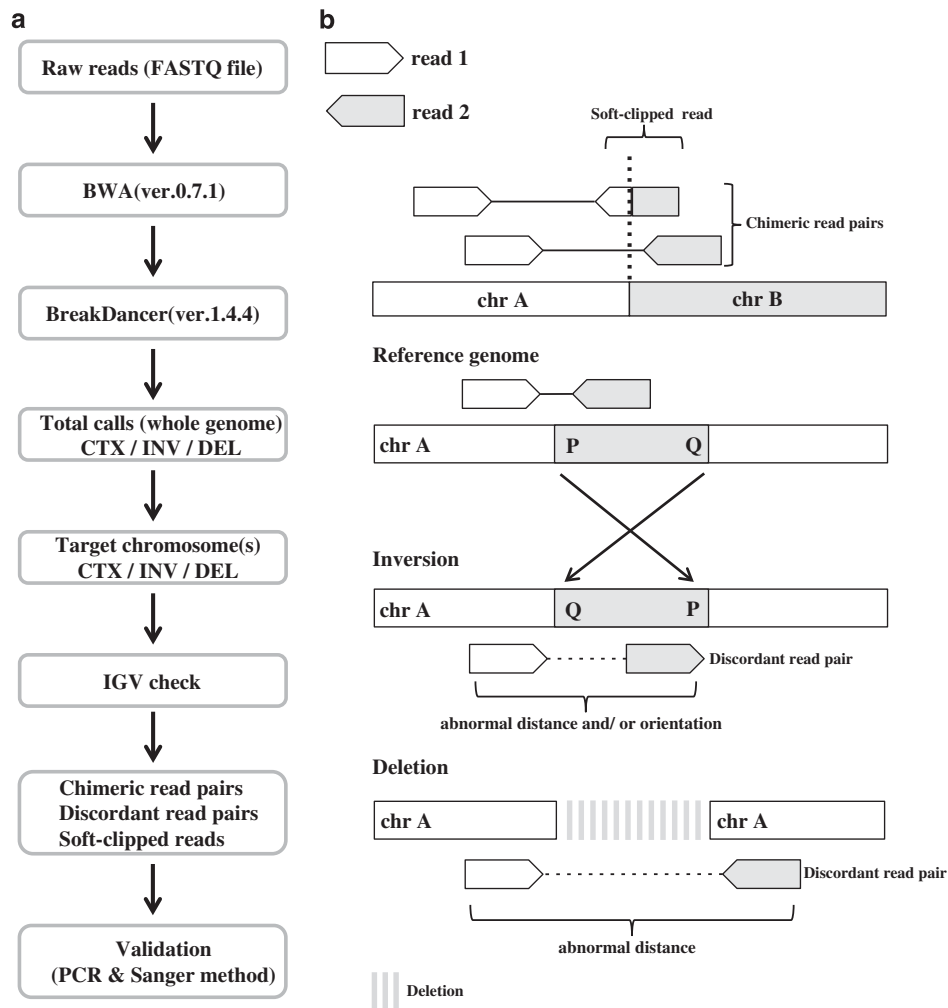
Figure 1 Schematic presentation of the analytical flow and graphical presentation of chimeric read pairs, discordant read pairs and a soft-clipped read. (**a**) Flowchart of structural variation (SV) breakpoint analysis using whole-genome sequencing (WGS) data. Burrows-Wheeler Aligner (BWA-MEM) v0.7.1 was used to map WGS data to human genome hg19. Inter-chromosomal translocation (CTX), inversion (INV) and deletion (DEL) were predicted using BreakDancerMax (BD). CTX, INV and DEL calls were selected only from the involved chromosomes. Furthermore, we focused on chimeric read pairs, discordant read pairs and soft-clipped reads using Integrative Genomics Viewer (IGV). Finally, we confirmed SV breakpoints by Sanger sequencing. (**b**) Graphical images of chimeric read pairs, discordant read pairs and a soft-clipped read. Upper, illustration of chimeric read pairs covering a translocation breakpoint, t(A;B). A soft-clipped read covers the breakpoint. Middle, intrachromosomal inversion. P and Q are marked to show the orientation. Inversion may lead to discordant read pairs. Lower, discordant read pairs detect intrachromosomal deletion.

SV breakpoint analysis

The analytical flowchart is illustrated in Figure 1a. Burrows-Wheeler Aligner (BWA-MEM) v0.7.1[15] with default parameters was used to map the data to the hg19 human genome reference sequence from the UCSC Genome Browser. BreakDancerMax (BD) ver.1.4.4 with the default setting was used to validate breakpoints of SVs, including translocations, inversions and deletions at the nucleotide level using the WGS data (Binary Alignment/Map format). A Poisson model[16] was used to calculate the confidence score for each candidate variant. BD is able to identify inter-chromosomal translocation (CTX), inversion (INV) and deletion (DEL). We focused on variant reads adjacent

to chromosomal breakpoint positions from the information of G-banded karyotyping. Aligned reads adjacent to SV breakpoints were visualized and carefully evaluated using Integrative Genomics Viewer (IGV).[17] In IGV, chimeric read pairs that mapped to different chromosomes at each end were predicted to cover translocation breakpoints (Figure 1b). Discordant read pairs that mapped to the reference genome with abnormal distance and/or orientation were predicted to cover breakpoints of inversion and insertion or deletion. Soft-clipped reads consisting of two different sequences (within a single read) mapped to discontinuous parts of chromosome(s) that potentially covered SV breakpoints (Figure 1b).

### Validation of chromosomal breakpoint positions
PCR and Sanger sequencing confirmed all potential SV breakpoints. Primer3-Plus (http://primer3plus.com/) was used to design the primer sequences. PCR was performed using KOD FX Neo polymerase (TOYOBO, Osaka, Japan). Primer sequences and PCR conditions are available on request. PCR products were electrophoresed through a 1.0% agarose gel and sequenced by Sanger sequencing on an ABI3500xl sequencer (Applied Biosystems, Foster City, CA, USA).

RESULTS
We analyzed 28 SVs in 9 patients. The analytical workflow of the respective patients is shown in Figure 2. Mean read depth of WGS was in the range of 5.95–21.92× (Table 1). Initially, the genomic DNA of each patient was sequenced using TruSeq DNA Sample Prep Kit. However, the read coverage did not reach the expected level because of high PCR duplication rates (Supplementary Table 1). Therefore, we switched the kit to the PCR-Free Sample Prep Kit and successfully attained the expected read coverage (Supplementary Table 1). We were able to detect 18 SV breakpoints of 28 (64.3%) using BD (Tables 1 and 2, and Figure 2).

For translocations and an inversion, the numbers of CTX and INV read by BD through a whole genome had a range of 61–4698 (Supplementary Table 2). We then focused on those related to the involved chromosome(s) by translocation or inversion, and found that 1–39 CTX or INV reads remained as candidates (Supplementary Table 2). Among the data, 12–31 chimeric read pairs and 28 discordant read pairs were carefully evaluated, which may have spanned SV breakpoints by IGV in patients 2, 3, 4, 6, 7, 8 and 9 (Table 1, Figure 3, and Supplementary Figure 1). In patients 1, 3, 4, 7, 8 and 9, one to six soft-clipped reads in IGV covered the SV breakpoints accurately (Table 1, Figure 3, and Supplementary Table 3). In patient 1, one CTX read by BD was a false positive (Supplementary Table 2); however, one soft-clipped read by IGV covered the 1q32 breakpoint (Table 1, Figures 2 and 3 and Supplementary Table 3). The 9q13 breakpoint region was undetected by either BD or IGV (Table 1, and Figures 2 and 3), because the genomic sequences of the region around centromeric 9q13 are unavailable. In combination with BD and IGV, 16 out of 18 translocations (88.9%) and 2 out of 2 inversion breakpoints (100%) were successfully determined.

Deletions in patient 3 (a 4192-bp deletion in the X-chromosome and a 7029-bp deletion in chromosome 4) and patient 4 (a 806 297-bp deletion in chromosome 7 and an ~4.6-Mb deletion in chromosome 15) were determined previously by conventional methods.[12,14] A total of 1943–1945 DEL reads were called by BD and 51–159 DEL reads related to the involved chromosomes remained as candidates; however, only one DEL read accurately covered the deletion breakpoint in chromosome 7. Therefore, we were able to detect the deletion breakpoints by BD in two of eight deletion breakpoints (25%) (Table 2 and Supplementary Table 4). Using IGV, nine discordant
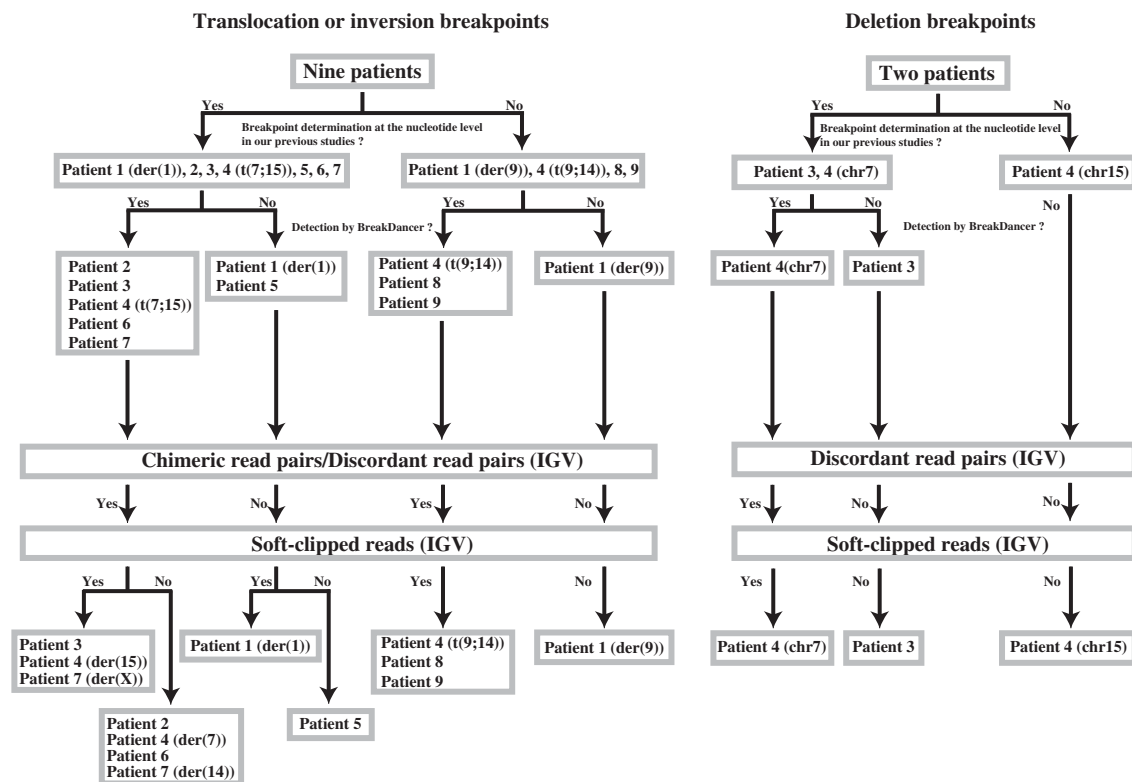
Figure 2 Flowchart of the analysis in nine patients. Left and right panels are shown in the detection of translocation/inversion and deletion breakpoints, respectively. BreakDancerMax (BD) was first used to detect structural variation (SV) breakpoints. We then carefully evaluated the chimeric read pairs, discordant read pairs and soft-clipped reads using Integrative Genomics Viewer (IGV). Validation was performed by the Sanger method.
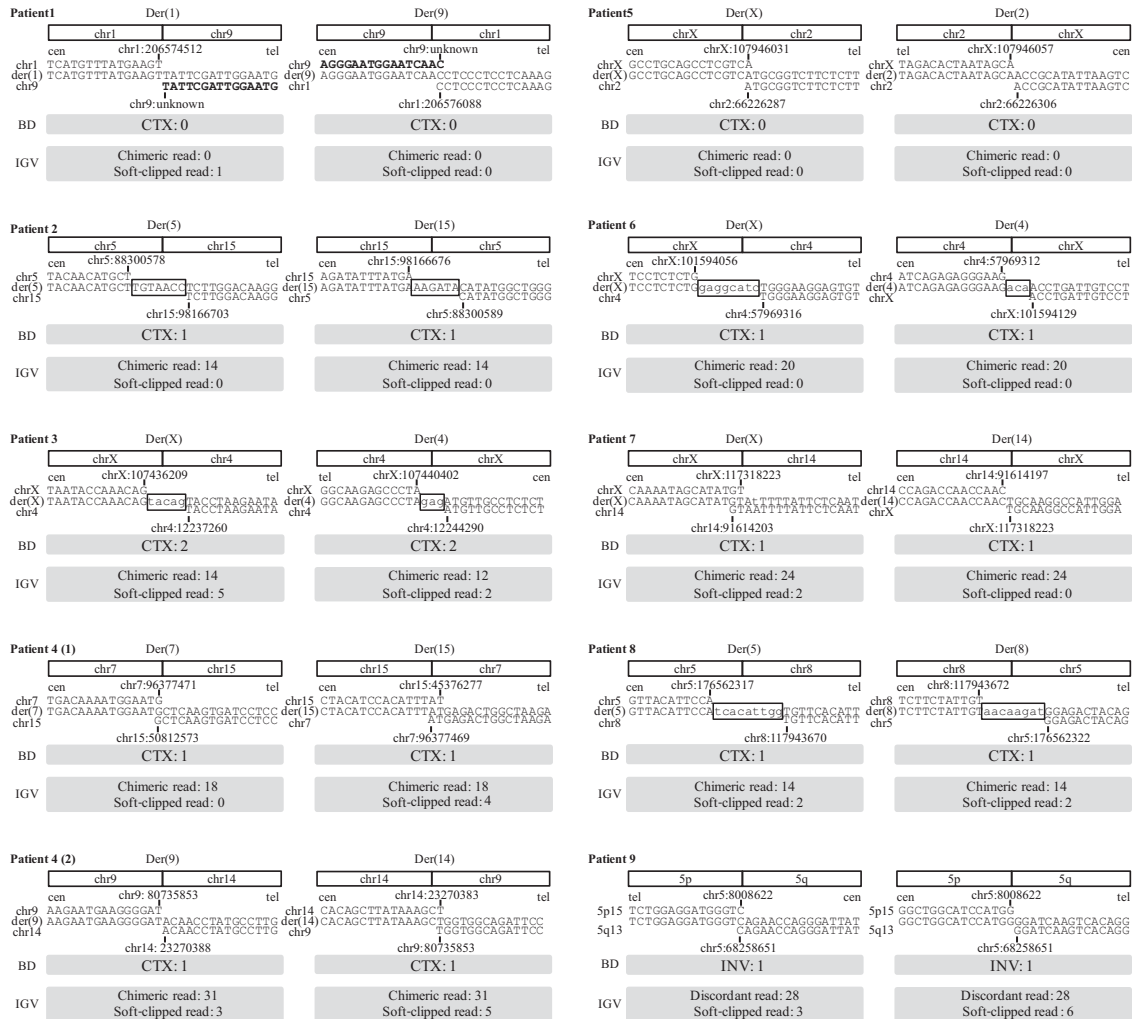
Figure 3 Breakpoint junction sequences in nine patients. Upper, middle and lower sequences indicate reference sequences of one end of a structural variation (SV), derivative/deleted chromosome, and the other end of the SV. Breakpoint positions are marked with short longitudinal lines. Numbers are based on the nucleotide position in the UCSC genome browser coordinates, February 2009 version (hg19). Bold sequences are novel sequences that have never been deposited to any databases. Boxes indicate nucleotide insertions. Total numbers of chimeric read pairs and soft-clipped reads are described. cen: centromere, tel: telomere.

read pairs accurately covered the deletion breakpoints in patient 4 (Table 2). Furthermore, two soft-clipped reads in IGV accurately covered the breakpoint in patient 4 (Table 2 and Supplementary Table 3). Of note, in patient 3, the deletions were adjacent to translocation breakpoints (Supplementary Figure 2).

DISCUSSION
In this study, 20 out of 28 SV breakpoints were successfully determined by WGS (71.4%). A relatively shallow (5×) read coverage enabled us to determine the translocation breakpoints (Table 1). Translocation and inversion breakpoints were highly detected by our method (88.9–100%), although the detection rate of deletion breakpoints was relatively low (25%). The false-negative rates by BD solely and BD combined with IGV were 10 out of 28 (35.7%) and 8 out of 28 (28.6%), respectively. The total number of called reads by BD including CTX, INV and DEL were quite different among samples (61–4698) (Supplementary Table 2 and 4). The estimation of the false positive rate (FPR) was difficult, because large and varying numbers of reads were called by BD. Therefore, FPR was unknown in the present study.

In patient 1, it was expected to be difficult or impossible to determine the 9q13 breakpoint because the genomic sequence data of the 9q13 centromeric region were unavailable. However, although no chimeric read pairs covering the breakpoints were obtained, one soft-clipped read accurately determined the der(1) breakpoint at the nucleotide level (Figures 2 and 3, Supplementary Table 3). The sequence with unknown origin in the soft-clipped read should be derived from the centromeric region at 9q13, as shown in a previous study.[10]

In patient 3, two CTX reads were called by BD (Supplementary Table 2). Interestingly, deletions existed adjacent to the reciprocal translocation in both chromosomes X and 4 (Supplementary Figure 2). However, BD did not call any DEL presumably because the sequences on either side of the deletion breakpoints are connected to different chromosomes.

In patients 4 (for t(9;14)), 8 and 9, translocation or inversion breakpoints had not been determined previously at the nucleotide level by any conventional method. We were able to determine the breakpoint positions of these patients by BD (Table 1 and Figures 2 and 3). A total of 14–31 chimeric read pairs or 28 discordant read pairs covered the SV breakpoints (Table 1 and Figure 3). Among

the soft-clipped reads, 2–6 reads also covered the precise breakpoints, including eight- or nine-nucleotide insertions of unknown origin (Table 1, and Supplementary Table 3).

In patient 5, chromosomal breakpoints could not be detected by our method (Table 1 and Supplementary Table 2). Breakpoint sequences were determined in the previous study, and no repetitive sequences and structural abnormalities were found around the breakpoints regardless of the relatively reasonable read coverage at the breakpoints (17 reads or 22 reads at Xq22.3 and 2p14, respectively). The reason for detection failure remains elusive.

The reason for the low detection rate of deletion breakpoints is that BD can only detect deletions with the sizes of <1 Mb. One 4.6-Mb deletion in which we were unable to determine deletion breakpoints was far beyond the size of the detection limit of BD. In addition, two deletions were adjacent to the translocation breakpoints in patient 3. Therefore, the two deletions were complicated. Each end of the two deletions and two translocation breakpoints are in the same location in patient 3. The only deletion in which we could determine breakpoints was the only simple 806-kb deletion within a single chromosome.

In conclusion, our approach, using shallow to moderate WGS data, enabled us to determine accurately the breakpoints of SVs, especially for chromosomal translocations and inversions. Conventional karyotyping, as well as the approximate localization of the SV breakpoints by FISH, was absolutely important for our WGS-based breakpoint detection. WGS analysis should be first considered for the determination of SV breakpoints in the NGS era.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

1   Feuk, L., Carson, A. R., Scherer, S. W. Structural variation in the human genome. Nat. Rev. Genet. 7, 85–97 (2006).
2   Utami, K. H., Hillmer, A. M., Aksoy, I., Chew, E. G., Teo, A. S., Zhang, Z. et al. Detection of chromosomal breakpoints in patients with developmental delay and speech disorders. PLoS One 9, e90852 (2014).
3   Vandeweyer, G., Kooy, R.F. Balanced translocations in mental retardation. Hum. Genet. 126, 133–147 (2009).
4   Fantes, J. A., Boland, E., Ramsay, J., Donnai, D., Splitt, M., Goodship, J. A. et al. FISH mapping of de novo apparently balanced chromosome rearrangements identifies characteristics associated with phenotypic abnormality. Am. J. Hum. Genet. 82, 916–926 (2008).
5   Rizzolio, F., Bione, S., Sala, C., Goegan, M., Gentile, M., Gregato, G. et al. Chromosomal rearrangements in Xq and premature ovarian failure: mapping of 25 new cases and review of the literature. Hum. Reprod. 21, 1477–1483 (2006).
6   Imaizumi, K., Kimura, J., Matsuo, M., Kurosawa, K., Masuno, M., Niikawa, N. et al. Sotos syndrome associated with a de novo balanced reciprocal translocation t(5;8)(q35;q24.1). Am. J. Med. Genet. 107, 58–60 (2002).
7   Dong, Z., Jiang, L., Yang, C., Hu, H., Wang, X., Chen, H. et al. A robust approach for blind detection of balanced chromosomal rearrangements with whole-genome low-coverage sequencing. Hum. Mutat. 35, 625–636 (2014).
8   Schluth-Bolard, C., Labalme, A., Cordier, M. P., Till, M., Nadeau, G., Tevissen, H. et al. Breakpoint mapping by next generation sequencing reveals causative gene disruption in patients carrying apparently balanced chromosome rearrangements with intellectual deficiency and/or congenital malformations. J. Med. Genet. 50, 144–150 (2013).
9   Abel, H. J., Duncavage, E. J. Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. Cancer Genet. 206, 432–440 (2013).
10 Saitsu, H., Osaka, H., Sugiyama, S., Kurosawa, K., Mizuguchi, T., Nishiyama, K. et al. Early infantile epileptic encephalopathy associated with the disrupted gene encoding Slit-Robo Rho GTPase activating protein 2 (SRGAP2). Am. J. Med. Genet. A 158A, 199–205 (2012).
11 Saitsu, H., Igarashi, N., Kato, M., Okada, I., Kosho, T., Shimokawa, O. et al. De novo 5q14.3 translocation 121.5-kb upstream of MEF2C in a patient with severe intellectual disability and early-onset epileptic encephalopathy. Am. J. Med. Genet. A 155A, 2879–2884 (2011).
12 Nishimura-Tadaki, A., Wada, T., Bano, G., Gough, K., Warner, J., Kosho, T. et al. Breakpoint determination of X;autosome balanced translocations in four patients with premature ovarian failure. J. Hum. Genet. 56, 156–160 (2011).
13 Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T. et al. Haploinsufficiency of NSD1 causes Sotos syndrome. Nat. Genet. 30, 365–366 (2002).
14 Saitsu, H., Kurosawa, K., Kawara, H., Eguchi, M., Mizuguchi, T., Harada, N. et al. Characterization of the complex 7q21.3 rearrangement in a patient with bilateral split-foot malformation and hearing loss. Am. J. Med. Genet. A 149A, 1224–1230 (2009).
15 Li, H., Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760 (2009).
16 Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S. et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat. Methods 6, 677–681 (2009).
17 Thorvaldsdottir, H., Robinson, J.T., Mesirov, J.P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. 14, 178–192 (2013).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg)